MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

UNITED STATES: THE END OF PHARMACOECONOMICS IN PHARMACY AS A SUBJECT WITHOUT MEASUREMENT CREDIBILITY

Paul C Langley Ph.D Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 5 JANUARY 2026

www.maimonresearch.com

Tucson AZ

FOREWORD

HEALTH TECHNOLOGY ASSESSMENT: A GLOBAL SYSTEM OF NON-MEASUREMENT

This Logit Working Paper series documents a finding as extraordinary as it is uncomfortable: health technology assessment (HTA), across nations, agencies, journals, and decades, has developed as a global system of non-measurement. It speaks the language of numbers, models, utilities, QALYs, "value for money," thresholds, discounting, incremental ratios, extrapolations, and simulations. It demands arithmetic at every turn, multiplication, division, summation, aggregation, discounting, yet it never once established that the quantities to which these operations are applied are measurable. HTA has built a vast evaluative machinery on foundations that do not exist. The probabilities and normalized logits in the reports that follow provide the empirical confirmation of this claim. They show, with unsettling consistency, that the global HTA knowledge base neither possesses nor applies the principles of scientific measurement.

The objective of this study is to assess, using a standardized 24-statement diagnostic instrument, the extent to which the teaching of pharmacoeconomics and health technology assessment in U.S. schools and colleges of pharmacy recognizes and applies the axioms of representational measurement theory. The analysis treats pharmacy education as a formative epistemic environment in which beliefs about what constitutes legitimate evidence, admissible arithmetic, and defensible decision rules are established and normalized. The study asks whether measurement validity is taught as a prerequisite for arithmetic operations such as multiplication, aggregation, and ratio construction, or whether arithmetic is presented as autonomous from measurement.

A secondary objective is explicitly professional. Pharmacy graduates are increasingly expected to serve as formulary managers, HTA analysts, and members of decision-making committees responsible for access, pricing, and utilization policy. The study therefore examines whether the educational knowledge base equips graduates to defend quantitative claims under scientific scrutiny, or whether it embeds a belief system that exposes them to professional, ethical, and credibility risk as challenges to QALYs, utilities, and modeling intensify.

The findings show that the teaching of pharmacoeconomics and HTA in U.S. schools and colleges of pharmacy systematically excludes the axioms of representational measurement while strongly endorsing the numerical propositions required to sustain cost-utility analysis, QALYs, and reference-case simulation modeling. Fundamental requirements, such as the necessity that measurement precede arithmetic, that multiplication requires ratio measures, and that arithmetic be licensed by scale properties are rejected or ignored. In contrast, false propositions essential to the prevailing curriculum, including that QALYs are ratio measures, that they are dimensionally homogeneous, and that they can be aggregated across individuals are strongly reinforced.

The result is an educational knowledge base in which arithmetic competence is treated as methodological rigor and measurement is treated as irrelevant. This inversion explains the routine acceptance of summated ordinal scores as ratio measures, the exclusion of Rasch measurement, and the presentation of simulation outputs as decision-relevant evidence. Pharmacy education does not merely transmit these beliefs; it institutionalizes them as professional norms.

The starting point is simple and inescapable: *measurement precedes arithmetic*. This principle is not a methodological preference but a logical necessity. One cannot multiply what one has not measured, cannot sum what has no dimensional homogeneity, cannot compare ratios when no ratio scale exists. When HTA multiplies time by utilities to generate QALYs, it is performing arithmetic with numbers that cannot support the operation. When HTA divides cost by QALYs, it is constructing a ratio from quantities that have no ratio properties. When HTA aggregates QALYs across individuals or conditions, it is combining values that do not share a common scale. These practices are not merely suboptimal; they are mathematically impossible.

The modern articulation of this principle can be traced to Stevens' seminal 1946 paper, which introduced the typology of nominal, ordinal, interval, and ratio scales [1]. Stevens made explicit what physicists, engineers, and psychologists already understood: different kinds of numbers permit different kinds of arithmetic. Ordinal scales allow ranking but not addition; interval scales permit addition and subtraction but not multiplication; ratio scales alone support multiplication, division, and the construction of meaningful ratios. Utilities derived from multiattribute preference exercises, such as EQ-5D or HUI, are ordinal preference scores; they do not satisfy the axioms of interval measurement, much less ratio measurement. Yet HTA has, for forty years, treated these utilities as if they were ratio quantities, multiplying them by time to create QALYs and inserting them into models without the slightest recognition that scale properties matter. Stevens' paper should have blocked the development of QALYs and cost-utility analysis entirely. Instead, it was ignored.

The foundational theory that establishes *when* and *whether* a set of numbers can be interpreted as measurements came with the publication of Krantz, Luce, Suppes, and Tversky's *Foundations of Measurement* (1971) [2] . Representational Measurement Theory (RMT) formalized the axioms under which empirical attributes can be mapped to numbers in a way that preserves structure. Measurement, in this framework, is not an act of assigning numbers for convenience, it is the discovery of a lawful relationship between empirical relations and numerical relations. The axioms of additive conjoint measurement, homogeneity, order, and invariance specify exactly when interval scales exist. RMT demonstrated once and for all that measurement is not optional and not a matter of taste: either the axioms hold and measurement is possible, or the axioms fail and measurement is impossible. Every major construct in HTA, utilities, QALYs, DALYs, ICERs, incremental ratios, preference weights, health-state indices, fails these axioms. They lack unidimensionality; they violate independence; they depend on aggregation of heterogeneous attributes; they collapse under the requirements of additive conjoint measurement. Yet HTA proceeded, decade after decade, without any engagement with these axioms, as if the field had collectively decided that measurement theory applied everywhere except in the evaluation of therapies.

Whereas representational measurement theory articulates the axioms for interval measurement, Georg Rasch's 1960 model provides the only scientific method for transforming ordered categorical responses into interval measures for latent traits [3]. Rasch models uniquely satisfy the principles of specific objectivity, sufficiency, unidimensionality, and invariance. For any construct such as pain, fatigue, depression, mobility, or need, Rasch analysis is the only legitimate means of producing an interval scale from ordinal item responses. Rasch measurement is not an alternative to RMT; it is its operational instantiation. The equivalence of Rasch's axioms and the axioms of representational measurement was demonstrated by Wright, Andrich and others as early as the 1970s. In the latent-trait domain, the very domain where HTA claims to operate; Rasch is the only game in town [4].

Yet Rasch is effectively absent from all HTA guidelines, including NICE, PBAC, CADTH, ICER, SMC, and PHARMAC. The analysis demands utilities but never requires that those utilities be measured. They rely on multiattribute ordinal classifications but never understand that those constructs be calibrated on interval or ratio scales. They mandate cost-utility analysis but never justify the arithmetic. They demand modelled QALYs but never interrogate their dimensional properties. These guidelines do not misunderstand Rasch; they do not know it exists. The axioms that define measurement and the model that makes latent trait measurement possible are invisible to the authors of global HTA rules. The field has evolved without the science that measurement demands.

How did HTA miss the bus so thoroughly? The answer lies in its historical origins. In the late 1970s and early 1980s, HTA emerged not from measurement science but from welfare economics, decision theory, and administrative pressure to control drug budgets. Its core concern was *valuing health states*, not *measuring health*. This move, quiet, subtle, but devastating, shifted the field away from the scientific question "What is the empirical structure of the construct we intend to measure?" and toward the administrative question "How do we elicit a preference weight that we can multiply by time?" The preference-elicitation projects of that era (SG, TTO, VAS) were rationalized as measurement techniques, but they never satisfied measurement axioms. Ordinal preferences were dressed up as quasi-cardinal indices; valuation tasks were misinterpreted as psychometrics; analyst convenience replaced measurement theory. The HTA community built an entire belief system around the illusion that valuing health is equivalent to measuring health. It is not.

The endurance of this belief system, forty years strong and globally uniform, is not evidence of validity but evidence of institutionalized error. HTA has operated under conditions of what can only be described as *structural epistemic closure*: a system that has never questioned its constructs because it never learned the language required to ask the questions. Representational measurement theory is not taught in graduate HTA programs; Rasch modelling is not part of guideline development; dimensional analysis is not part of methodological review. The field has been insulated from correction because its conceptual foundations were never laid. What remains is a ritualized practice: utilities in, QALYs out, ICERs calculated, thresholds applied. The arithmetic continues because everyone assumes someone else validated the numbers.

This Logit Working Paper series exposes, through probabilistic and logit-based interrogations of AI large language national knowledge bases, the scale of this failure. The results display a global pattern: true statements reflecting the axioms of measurement receive weak endorsement; false statements reflecting the HTA belief system receive moderate or strong reinforcement. This is not disagreement. It is non-possession. It shows that HTA, worldwide, has developed as a quantitative discipline without quantitative foundations; a confused exercise in numerical storytelling.

The conclusion is unavoidable: HTA does not need incremental reform; it needs a scientific revolution. Measurement must precede arithmetic. Representational axioms must precede valuation rituals. Rasch measurement must replace ordinal summation and utility algorithms. Value claims must be falsifiable, protocol-driven, and measurable; rather than simulated, aggregated, and numerically embellished.

The global system of non-measurement is now visible. The task ahead is to replace it with science.

Paul C Langley, Ph.D

Email: langleylapaloma@gmail.com

DISCLAIMER

This analysis is generated through the structured interrogation of a large language model (LLM) applied to a defined documentary corpus and is intended solely to characterize patterns within an aggregated knowledge environment. It does identify, assess, or attribute beliefs, intentions, competencies, or actions to any named individual, faculty member, student, administrator, institution, or organization. The results do not constitute factual findings about specific persons or programs, nor should they be interpreted as claims regarding professional conduct, educational quality, or compliance with regulatory or accreditation standards. All probabilities and logit values reflect model-based inferences about the presence or absence of concepts within a bounded textual ecosystem, not judgments about real-world actors. The analysis is exploratory, interpretive, and methodological in nature, offered for scholarly discussion of epistemic structures rather than evaluative or legal purposes. Any resemblance to particular institutions or practices is contextual and non-attributive, and no adverse implication should be inferred.

1. INTERROGATING THE LARGE LANGUAGE MODEL

A large language model **(**LLM**)** is an artificial intelligence system designed to understand, generate, and manipulate human language by learning patterns from vast amounts of text data. Built on deep neural network architectures, most commonly transformers, LLMs analyze relationships between words, sentences, and concepts to produce contextually relevant responses. During training, the model processes billions of examples, enabling it to learn grammar, facts, reasoning patterns, and even subtle linguistic nuances. Once trained, an LLM can perform a wide range of tasks: answering questions, summarizing documents, generating creative writing, translating languages, assisting with coding, and more. Although LLMs do not possess consciousness or true understanding, they simulate comprehension by predicting the most likely continuation of text based on learned patterns. Their capabilities make them powerful tools for communication, research, automation, and decision support, but they also require careful oversight to ensure accuracy, fairness, privacy, and responsible use

In this Logit Working Paper, "interrogation" refers not to discovering what an LLM *believes,* it has no beliefs, but to probing the content of the *corpus-defined knowledge space* we choose to analyze. This knowledge base is enhanced if it is backed by accumulated memory from the user. In this case the interrogation relies also on 12 months of HTA memory from continued application of the system to evaluate HTA experience. The corpus is defined before interrogation: it may consist of a journal (e.g., *Value in Health*), a national HTA body, a specific methodological framework, or a collection of policy documents. Once the boundaries of that corpus are established, the LLM is used to estimate the conceptual footprint within it. This approach allows us to determine which principles are articulated, neglected, misunderstood, or systematically reinforced.

In this HTA assessment, the objective is precise: to determine the extent to which a given HTA knowledge base or corpus, global, national, institutional, or journal-specific, recognizes and reinforces the foundational principles of representational measurement theory (RMT). The core principle under investigation is that measurement precedes arithmetic; no construct may be treated as a number or subjected to mathematical operations unless the axioms of measurement are satisfied. These axioms include unidimensionality, scale-type distinctions, invariance, additivity, and the requirement that ordinal responses cannot lawfully be transformed into interval or ratio quantities except under Rasch measurement rules.

The HTA knowledge space is defined pragmatically and operationally. For each jurisdiction, organization, or journal, the corpus consists of:

- published HTA guidelines
- agency decision frameworks
- cost-effectiveness reference cases
- academic journals and textbooks associated with HTA
- modelling templates, technical reports, and task-force recommendations
- teaching materials, methodological articles, and institutional white papers

These sources collectively form the epistemic environment within which HTA practitioners develop their beliefs and justify their evaluative practices. The boundary of interrogation is thus not the whole of medicine, economics, or public policy, but the specific textual ecosystem that sustains HTA reasoning. . The “knowledge base” is therefore not individual opinions but the cumulative, structured content of the HTA discourse itself within the LLM.

PHARMACY HTA EDUCATION AND PRACTICE ASSESSMENT KNOWLEDGE BASE

The knowledge base interrogated in this study is defined as the cumulative, structured content that shapes how pharmacoeconomics and HTA are taught, learned, and legitimized within U.S. colleges and schools of pharmacy. It is not a survey of personal beliefs, nor an assessment of individual competence. Rather, it represents the textual and institutional environment through which epistemic norms are transmitted to successive cohorts of pharmacists.

The corpus includes required and recommended pharmacoeconomics and outcomes research textbooks used in PharmD programs; ACPE-aligned curricular materials and competency statements; syllabi and lecture content from core courses in pharmacoeconomics, outcomes research, managed care, and HTA; continuing education materials endorsed by pharmacy organizations; and peer-reviewed articles routinely assigned or cited in pharmacy training. It also incorporates methodological frameworks adopted implicitly or explicitly in teaching, including cost-utility analysis, incremental cost-effectiveness ratios, reference-case modeling conventions, and the routine use of utilities, QALYs, and preference-based instruments.

In addition, the knowledge base encompasses the professional signaling environment surrounding pharmacy education: examination expectations, credentialing pathways, fellowship preparation materials, and institutional affiliations with organizations such as ISPOR, AMCP, and managed-care bodies. These sources collectively define what is treated as legitimate knowledge, acceptable method, and professional common sense within pharmacy education.

Crucially, the boundary of this knowledge base excludes disciplines that would otherwise correct its errors. Formal training in measurement theory, representational measurement axioms, Rasch modeling, or scale-type analysis is largely absent from pharmacy curricula and is therefore not part of the environment being interrogated. The analysis asks a precise question: given what pharmacy students are actually taught and rewarded for mastering, what conception of measurement does this system possess?

The answer, as revealed by the logit profile, is that the pharmacy education knowledge base does not merely omit measurement theory, it actively operates as if it were unnecessary. This makes it an appropriate and consequential target for interrogation, because it is here that professional habits of evaluation are first formed and normalized.

CATEGORICAL PROBABILITIES

In the present application, the interrogation is tightly bounded. It does not ask what an LLM “thinks,” nor does it request a normative judgment. Instead, the LLM evaluates how likely the

HTA knowledge space is to endorse, imply, or reinforce a set of 24 diagnostic statements derived from representational measurement theory (RMT). Each statement is objectively TRUE or FALSE under RMT. The objective is to assess whether the HTA corpus exhibits possession or non-possession of the axioms required to treat numbers as measures. The interrogation creates a categorical endorsement probability: the estimated likelihood that the HTA knowledge base endorses the statement whether it is true or false; *explicitly or implicitly.*

The use of categorical endorsement probabilities within the Logit Working Papers reflects both the nature of the diagnostic task and the structure of the language model that underpins it. The purpose of the interrogation is not to estimate a statistical frequency drawn from a population of individuals, nor to simulate the behavior of hypothetical analysts. Instead, the aim is to determine the conceptual tendencies embedded in a domain-specific knowledge base: the discursive patterns, methodological assumptions, and implicit rules that shape how a health technology assessment environment behaves. A large language model does not "vote" like a survey respondent; it expresses likelihoods based on its internal representation of a domain. In this context, endorsement probabilities capture the strength with which the knowledge base, as represented within the model, supports a particular proposition. Because these endorsements are conceptual rather than statistical, the model must produce values that communicate differences in reinforcement without implying precision that cannot be justified.

This is why categorical probabilities are essential. Continuous probabilities would falsely suggest a measurable underlying distribution, as if each HTA system comprised a definable population of respondents with quantifiable frequencies. But large language models do not operate on that level. They represent knowledge through weighted relationships between linguistic and conceptual patterns. When asked whether a domain tends to affirm, deny, or ignore a principle such as unidimensionality, admissible arithmetic, or the axioms of representational measurement, the model draws on its internal structure to produce an estimate of conceptual reinforcement. The precision of that estimate must match the nature of the task. Categorical probabilities therefore provide a disciplined and interpretable way of capturing reinforcement strength while avoiding the illusion of statistical granularity.

The categories used, values such as 0.05, 0.10, 0.20, 0.50, 0.75, 0.80, and 0.85, are not arbitrary. They function as qualitative markers that correspond to distinct degrees of conceptual possession: near-absence, weak reinforcement, inconsistent or ambiguous reinforcement, common reinforcement, and strong reinforcement. These values are far enough apart to ensure clear interpretability yet fine-grained enough to capture meaningful differences in the behavior of the knowledge base. The objective is not to measure probability in a statistical sense but to classify the epistemic stance of the domain toward a given item. A probability of 0.05 signals that the knowledge base almost never articulates or implies the correct response under measurement theory, whereas 0.85 indicates that the domain routinely reinforces it. Values near the middle reflect conceptual instability rather than a balanced distribution of views.

Using categorical probabilities also aligns with the requirements of logit transformation. Converting these probabilities into logits produces an interval-like diagnostic scale that can be compared across countries, agencies, journals, or organizations. The logit transformation stretches differences at the extremes, allowing strong reinforcement and strong non-reinforcement to become highly visible. Normalizing logits to the fixed ±2.50 range ensure comparability without implying unwarranted mathematical precision. Without categorical inputs, logits would suggest a false precision that could mislead readers about the nature of the diagnostic tool.

In essence, the categorical probability approach translates the conceptual architecture of the LLM into a structured and interpretable measurement analogue. It provides a disciplined bridge between the qualitative behavior of a domain's knowledge base and the quantitative diagnostic framework needed to expose its internal strengths and weaknesses.

The LLM computes these categorical probabilities from three sources:

1. **Structural content of HTA discourse**
   If the literature repeatedly uses ordinal utilities as interval measures, multiplies non-quantities, aggregates QALYs, or treats simulations as falsifiable, the model infers high reinforcement of these false statements.
2. **Conceptual visibility of measurement axioms**
   If ideas such as unidimensionality, dimensional homogeneity, scale-type integrity, or Rasch transformation rarely appear, or are contradicted by practice, the model assigns low endorsement probabilities to TRUE statements.
3. **The model's learned representation of domain stability**
   Where discourse is fragmented, contradictory, or conceptually hollow, the model avoids assigning high probabilities. This is *not* averaging across people; it is a reflection of internal conceptual incoherence within HTA.

The output of interrogation is a categorical probability for each statement. Probabilities are then transformed into logits [ln(p/(1–p)], capped to ±4.0 logits to avoid extreme distortions, and normalized to ±2.50 logits for comparability across countries. A positive normalized logit indicates reinforcement in the knowledge base. A negative logit indicates weak reinforcement or conceptual absence. Values near zero logits reflect epistemic noise.

Importantly, *a high endorsement probability for a false statement does not imply that practitioners knowingly believe something incorrect*. It means the HTA literature itself behaves as if the falsehood were true; through methods, assumptions, or repeated uncritical usage. Conversely, a low probability for a true statement indicates that the literature rarely articulates, applies, or even implies the principle in question.

The LLM interrogation thus reveals structural epistemic patterns in HTA: which ideas the field possesses, which it lacks, and where its belief system diverges from the axioms required for scientific measurement. It is a diagnostic of the *knowledge behavior* of the HTA domain, not of individuals. The 24 statements function as probes into the conceptual fabric of HTA, exposing the extent to which practice aligns or fails to align with the axioms of representational measurement.

INTERROGATION STATEMENTS

Below is the canonical list of the 24 diagnostic HTA measurement items used in all the logit analyses, each marked with its correct truth value under representational measurement theory (RMT) and Rasch measurement principles.

This is the definitive set used across the Logit Working Papers.

Measurement Theory & Scale Properties

1. Interval measures lack a true zero — TRUE
2. Measures must be unidimensional — TRUE
3. Multiplication requires a ratio measure — TRUE
4. Time trade-off preferences are unidimensional — FALSE
5. Ratio measures can have negative values — FALSE
6. EQ-5D-3L preference algorithms create interval measures — FALSE
7. The QALY is a ratio measure — FALSE
8. Time is a ratio measure — TRUE

Measurement Preconditions for Arithmetic

9. Measurement precedes arithmetic — TRUE
10. Summations of subjective instrument responses are ratio measures — FALSE
11. Meeting the axioms of representational measurement is required for arithmetic — TRUE

Rasch Measurement & Latent Traits

12. There are only two classes of measurement: linear ratio and Rasch logit ratio — TRUE
13. Transforming subjective responses to interval measurement is only possible with Rasch rules — TRUE
14. Summation of Likert question scores creates a ratio measure — FALSE

Properties of QALYs & Utilities

15. The QALY is a dimensionally homogeneous measure — FALSE
16. Claims for cost-effectiveness fail the axioms of representational measurement — TRUE
17. QALYs can be aggregated — FALSE

Falsifiability & Scientific Standards

18. Non-falsifiable claims should be rejected — TRUE
19. Reference-case simulations generate falsifiable claims — FALSE

Logit Fundamentals

20. The logit is the natural logarithm of the odds-ratio — TRUE

Latent Trait Theory

21. The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits — TRUE
22. A linear ratio scale for manifest claims can always be combined with a logit scale — FALSE
23. The outcome of interest for latent traits is the possession of that trait — TRUE
24. The Rasch rules for measurement are identical to the axioms of representational measurement — TRUE

AI LARGE LANGUAGE MODEL STATEMENTS: TRUE OR FALSE

Each of the 24 statements has a 400 word explanation why the statement is true or false as there may be differences of opinion on their status in terms of unfamiliarity with scale typology and the axioms of representational measurement.

The link to these explanations is: https://maimonresearch.com/ai-llm-true-or-false/

INTERPRETING TRUE STATEMENTS

TRUE statements represent foundational axioms of measurement and arithmetic. Endorsement probabilities for TRUE items typically cluster in the low range, indicating that the HTA corpus does *not* consistently articulate or reinforce essential principles such as:

- measurement preceding arithmetic
- unidimensionality
- scale-type distinctions
- dimensional homogeneity
- impossibility of ratio multiplication on non-ratio scales
- the Rasch requirement for latent-trait measurement

Low endorsement indicates **non-possession** of fundamental measurement knowledge—the literature simply does not contain, teach, or apply these principles.

INTERPRETING FALSE STATEMENTS

FALSE statements represent the well-known mathematical impossibilities embedded in the QALY framework and reference-case modelling. Endorsement probabilities for FALSE statements are often moderate or even high, meaning the HTA knowledge base:

- accepts non-falsifiable simulation as evidence

- permits negative “ratio” measures
- treats ordinal utilities as interval measures
- treats QALYs as ratio measures
- treats summated ordinal scores as ratio scales
- accepts dimensional incoherence

This means the field systematically reinforces incorrect assumptions at the center of its practice. *Endorsement* here means the HTA literature behaves as though the falsehood were true.

2. SUMMARY OF FINDINGS FOR TRUE AND FALSE ENDORSEMENTS: PHARMACY TEACHING AND PRACTICE

Table 1 presents, the endorsement probabilities and normalized logits for each of the 24 diagnostic measurement statements. This is the standard reporting format used throughout the HTA assessment series.

It is essential to understand how to interpret these results.

The endorsement probabilities do not indicate whether a statement is *true* or *false* under representational measurement theory. Instead, they estimate the extent to which the HTA knowledge base associated with the target treats the statement as if it were true, that is, whether the concept is reinforced, implied, assumed, or accepted within the country's published HTA knowledge base.

The logits provide a continuous, symmetric scale, ranging from +2.50 to –2.50, that quantifies the degree of this endorsement. the logits, of course link to the probabilities (p) as the logit is the natural logarithm of the odds ratio; logit = ln[p/1-p].

- Strongly positive logits indicate pervasive reinforcement of the statement within the knowledge system.
- Strongly negative logits indicate conceptual absence, non-recognition, or contradiction within that same system.
- Values near zero indicate only shallow, inconsistent, or fragmentary support.

Thus, the endorsement logit profile serves as a direct index of a country or organization's epistemic alignment with the axioms of scientific measurement, revealing the internal structure of its HTA discourse. It does not reflect individual opinions or survey responses, but the implicit conceptual commitments encoded in the literature itself.

TABLE 1: ITEM STATEMENT, RESPONSE, ENDORSEMENT AND NORMALIZED PHARMACY TEACHING AND PRACTICE

STATEMENT	**RESPONSE 1=TRUE 0=FALSE**	**ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY**	**NORMALIZED LOGIT (IN RANGE +/- 2.50)**
INTERVAL MEASURES LACK A TRUE ZERO	1	0.20	-1.40
MEASURES MUST BE UNIDIMENSIONAL	1	0.20	-1.40

MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.20
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.80	+1.40
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.90	+2.20
EQ-5D-3L PREFERENCE ALGORITHMS CREATE INTERVAL MEASURES	0	0.90	+2.20
THE QALY IS A RATIO MEASURE	0	0.90	+2.20
TIME IS A RATIO MEASURE	1	0.95	+2.50
MEASUREMENT PRECEDES ARITHMETIC	1	0.10	-2.20
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.85	+1.75
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.10	-2.20
THERE ARE ONLY TWO CLASSES OF MEASUREMENT LINEAR RATIO AND RASCH LOGIT RATIO	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50
SUMMATION OF LIKERT QUESTION SCORES CREATES A RATIO MEASURE	0	0.90	+2.20
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.85	+1.75
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.15	-1.75
QALYS CAN BE AGGREGATED	0	0.95	+2.50
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.75	+1.10
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.90	+2.20
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.65	+0.60
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FOR ASSESSING THERAPY IMPACT FOR LATENT TRAITS	1	0.05	-2.50

A LINEAR RATIO SCALE FOR MANIFEST CLAIMS CAN ALWAYS BE COMBINED WITH A LOGIT SCALE	0	0.65	+0.60
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.20	-1.40
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.05	-2.50

THE END OF PHARMACOECONOMICS: A SUBJECT WITHOUT MEASUREMENT

For more than four decades, the teaching of pharmacoeconomics and health technology assessment in U.S. schools and colleges of pharmacy has presented itself as rigorous, quantitative, and scientifically grounded. Generations of students have been trained to believe that they are acquiring the analytical tools necessary to evaluate therapy impact, inform formulary decisions, and support evidence-based healthcare. Yet when this educational corpus is interrogated against the axioms of representational measurement theory, a stark conclusion emerges: pharmacy education has systematically failed to teach what measurement is, why it matters, and how it constrains arithmetic. The result is not a minor pedagogical gap but a foundational epistemic failure that has misled students, faculty, and health systems alike.

Measurement is not a technical afterthought. It is the logical prerequisite for any quantitative claim. Without measurement, numbers do not represent quantities; they are merely symbols manipulated by convention. Representational measurement theory makes this explicit. Arithmetic operations are only meaningful when the scale properties of the underlying measures permit them. Addition requires at least interval scales; multiplication and division require ratio scales with a true zero. These are not philosophical niceties. They are the conditions under which numerical claims can be falsified, replicated, and embedded in cumulative scientific knowledge. To teach quantitative evaluation while ignoring these conditions is to teach numerical storytelling rather than science.

The 24-item diagnostic applied to U.S. pharmacy education reveals that this is precisely what has occurred. Core axioms that should govern any quantitative curriculum are weakly endorsed or outright rejected. The proposition that measurement must precede arithmetic collapses to near-floor endorsement. The requirement that multiplication demands a ratio measure fares no better. Unidimensionality, the most basic condition for defining what is being measured, is treated as optional. These results are not accidental. They reflect what is routinely taught, reinforced, and assessed in pharmacoeconomics and HTA courses across the country.

At the same time, propositions that are mathematically impossible under representational measurement theory receive near-ceiling endorsement. Students are taught, explicitly or implicitly, that summing Likert-scale responses creates ratio measures, that utilities derived from preference instruments are ratio-scaled quantities, that negative values on supposed ratio scales are acceptable, and that QALYs can be aggregated across individuals and populations. These beliefs are not fringe errors. They are embedded in textbooks, course syllabi, model templates, and examination questions. They define what students are expected to know in order to pass, graduate, and be credentialed as competent analysts.

This combination—rejection of measurement axioms alongside strong endorsement of impossible arithmetic—defines a structural inversion. Arithmetic is treated as primary, authoritative, and self-justifying, while measurement is treated as an implicit assumption or ignored entirely. The inversion is pedagogically catastrophic. Students learn how to compute before they learn whether computation is legitimate. They are trained to manipulate ICERs, utilities, and QALYs without ever being taught to ask whether these objects satisfy the conditions required for arithmetic to be meaningful. By the time they reach professional practice, the inversion has hardened into habit.

The consequences extend far beyond the classroom. Pharmacy graduates populate managed care organizations, PBMs, hospital formularies, HTA agencies, and advisory committees. They carry with them the belief system instilled during training. When they encounter cost-effectiveness models, reference-case simulations, or utility-based claims, they rarely question their measurement foundations, not because the foundations are sound, but because they were never taught to interrogate them. The educational system has already normalized the idea that quantification does not require measurement discipline.

The most damaging aspect of this pedagogical failure is its treatment of subjective outcomes. Pharmacy education places heavy emphasis on patient-reported outcomes, quality of life instruments, and preference-based measures. Students are taught that these instruments capture "value" or "benefit" in a way that can be directly incorporated into economic evaluation. Yet the diagnostic shows near-total rejection of the only framework capable of transforming ordinal subjective responses into legitimate quantitative measures: Rasch measurement. The propositions linking latent traits to Rasch logit ratio scales collapse to the floor of the scale. This indicates not merely neglect but structural exclusion.

This exclusion has profound implications. Rasch measurement is not one psychometric option among many. It is the only model that satisfies the requirements of conjoint measurement for latent traits, producing invariant units that permit meaningful comparisons across persons and items. Without Rasch, subjective responses remain ordinal. They can rank individuals but cannot quantify how much of a trait is possessed. By rejecting Rasch while endorsing summation-based scoring, pharmacy education teaches students to treat ordinal scores as if they were quantities. This is not measurement. It is numerology.

The concept of possession of a latent trait, which lies at the heart of any defensible quality-of-life claim, is correspondingly marginalized. The diagnostic shows weak endorsement of the proposition that the outcome of interest for latent traits is possession of that trait. Instead, students are trained to focus on changes in scores, mean differences, responder thresholds, and minimally important differences, all computed on scales that lack interval or ratio properties. These operations give the illusion of precision while systematically evading the question of what, exactly, is being measured.

The QALY occupies a central role in this educational failure. Pharmacy students are routinely taught that QALYs represent a coherent, quantitative measure of health benefit. They are trained to calculate, compare, and aggregate QALYs as if they were ratio-scaled quantities. The diagnostic confirms that these beliefs are reinforced at near-ceiling levels. Yet under representational measurement theory, the QALY is indefensible. It multiplies time, a ratio measure, by utilities that are at best ordinal. It aggregates across individuals without demonstrating dimensional homogeneity. It permits negative values despite claiming ratio properties. None of these violations are marginal; they are fatal.

What makes the persistence of the QALY in pharmacy education particularly troubling is that it is taught not as a contested construct but as settled science. Students are rarely exposed to measurement critiques. Representational measurement theory is absent from curricula. Rasch measurement, if mentioned at all, is treated as an advanced or optional technique rather than as a governing requirement. As a result, students are deprived of the conceptual tools necessary to evaluate the validity of the very constructs they are told to rely on.

The teaching of modeling compounds the problem. Pharmacy students are trained to view reference-case simulation models as legitimate sources of evidence. Sensitivity analyses are presented as demonstrations of robustness. Probabilistic outputs are treated as if they confer falsifiability. The diagnostic shows strong endorsement of the false belief that reference-case simulations generate falsifiable claims. They do not. Models generate conditional projections based on assumptions. Exploring how outputs change when assumptions vary is not falsification; it is internal consistency checking. By conflating robustness with falsifiability, pharmacy education teaches students to mistake stability for truth.

This pedagogical environment is incompatible with the evolution of objective knowledge. Without measurement, claims cannot be refuted in the Popperian sense. Without invariant units, replication becomes impossible. Studies cannot accumulate into a coherent body of knowledge because there is no shared quantitative structure to reproduce. What evolves instead is convention. Methods persist because they are taught, not because they have survived empirical challenge. The educational system thus functions as a replication engine for a belief system rather than as a training ground for scientific inquiry.

The longevity of this failure—now exceeding forty years—demands explanation. It cannot be attributed to ignorance. Representational measurement theory has been well established since the mid-twentieth century. The Rasch model has been available for over sixty years. The issue is not lack of knowledge but lack of incentives. Teaching measurement axioms would destabilize the core content of pharmacoeconomics curricula. It would force educators to confront the invalidity of widely used tools, invalidate standard examination questions, and challenge professional identities built around cost-effectiveness analysis. The path of least resistance has been to ignore measurement constraints and proceed as if arithmetic alone conferred legitimacy.

The ethical implications are significant. Students trust that accredited programs are teaching them valid scientific methods. Health systems trust that graduates are equipped to evaluate evidence responsibly. Patients trust that access and pricing decisions are grounded in sound analysis. When education fails at the level of measurement, that trust is violated. Decisions are made on the basis of claims that cannot, in principle, be validated. Opportunity costs, access restrictions, and pricing benchmarks are justified using numbers that do not measure what they purport to measure.

Reform must begin with education. Measurement cannot be an elective topic or a footnote. It must be the foundation of any quantitative curriculum. Students must be taught, early and explicitly, that arithmetic is conditional on measurement, not the other way around. They must learn to distinguish ordinal from interval and ratio scales, to recognize the impossibility of certain operations, and to understand why unidimensionality is non-negotiable. They must be taught that subjective outcomes require Rasch measurement if they are to be quantified, and that without it, claims of “amount” or “change” are illegitimate.

Such reform would be disruptive. It would require abandoning the QALY as a teaching construct. It would require reframing cost-effectiveness analysis as a descriptive or heuristic exercise rather than as quantitative evidence. It would require redesigning assessments, textbooks, and accreditation standards. Yet disruption is precisely what scientific correction demands. Continuing to teach false measurement is not pragmatic; it is irresponsible.

The 24-item diagnostic leaves no room for complacency. The belief system embedded in U.S. pharmacy education is internally coherent but scientifically indefensible. It systematically endorses arithmetic without measurement, rejects the axioms that would constrain that arithmetic, and marginalizes the only models capable of rescuing subjective measurement. This is not a failure at the margins. It is a failure at the core of what is taught as quantitative reasoning.

If pharmacy education is to claim a role in evidence-based healthcare, it must confront this failure directly. Measurement must be restored to its rightful place as the gatekeeper of arithmetic and the foundation of falsifiable claims. Until that happens, students will continue to be misled, health systems will continue to rely on pseudo-quantitative evidence, and the evolution of objective knowledge in therapy assessment will remain stalled.

HOW WAS THIS EPISTEMIC DISASTER ACHIEVED

This epistemic disaster did not arise from a single error, nor from the incompetence of individual instructors. It was achieved through a long sequence of institutional choices that normalized measurement failure while presenting it as technical sophistication. The central mechanism was not deception, but omission: the systematic exclusion of representational measurement theory from pharmacy education, coupled with the uncritical adoption of HTA conventions as settled science.

The first step was curricular framing. Pharmacoeconomics and HTA entered pharmacy education as applied, policy-facing disciplines, positioned as tools for decision making rather than as sciences requiring foundational scrutiny. Students were taught how to calculate cost-effectiveness ratios, manipulate utilities, and populate reference-case models without ever being taught the prior question: *what conditions must be satisfied before numbers may lawfully be added, multiplied, or compared?* By design or neglect, arithmetic was taught in the absence of measurement. Once this inversion became routine, it ceased to be visible.

Second, pedagogy substituted technique for theory. Courses emphasized software use, model structure, discounting conventions, sensitivity analysis, and guideline compliance. These are all procedural skills. None require students to understand scale types, unidimensionality, invariance, or the axioms that distinguish ordinal scores from interval or ratio measures. As a result, graduates emerged fluent in methods while epistemically illiterate about the quantities those methods presuppose. The illusion of rigor was complete: complexity replaced validity.

Third, external validation reinforced the error. Accreditation standards, professional competencies, and recommended curricula echoed the same methodological assumptions found in HTA agencies, journals, and professional societies. Because these external bodies themselves operate within a measurement-deficient framework, pharmacy schools received no corrective signal. Teaching QALYs and ICERs was not merely acceptable; it was expected. No institution was penalized for teaching arithmetic on non-measures. On the contrary, alignment with NICE, ICER, CADTH, and ISPOR was treated as evidence of academic seriousness.

Fourth, the absence of disciplinary cross-pollination sealed the outcome. Representational measurement theory, Rasch measurement, and scale theory are not obscure or controversial bodies of knowledge. They are standard in psychology, education, and the measurement sciences. Yet pharmacy education remained closed, drawing methodological authority almost exclusively from health economics and HTA practice rather than from measurement science. This intellectual isolation ensured that students never encountered the critique that would have made the contradictions visible.

Fifth, professional incentives completed the process. Faculty who taught conventional pharmacoeconomics could publish, consult, advise agencies, and secure grants without ever addressing measurement validity. There was no reward for introducing measurement theory, and substantial risk in challenging accepted constructs such as utilities and QALYs. Over time, what began as unexamined convention hardened into orthodoxy. New faculty were socialized into the same framework, not because it was defended, but because it was ubiquitous.

The final step was normalization through repetition. Cohort after cohort of students was trained to accept that subjective preference scores are measures, that multiattribute health states can be reduced to single numbers, and that lifetime simulations constitute evidence. Because everyone learned the same thing, the possibility that it might be wrong never arose. The system achieved stability not through proof, but through uniformity.

In this way, the epistemic disaster was achieved quietly, efficiently, and without confrontation. No conspiracy was required. Only the consistent failure to teach that measurement precedes arithmetic, and the willingness of institutions to accept borrowed conventions as scientific fact. The result is a generation of pharmacy graduates equipped to produce cost-effectiveness analyses, but unequipped to recognize that the quantities on which those analyses rest do not exist as measures at all.

IMPLICATIONS FOR PHARMACISTS AND DUTY OF CARE

The epistemic failures embedded in the teaching of pharmacoeconomics and health technology assessment are not abstract academic concerns. They have direct and serious implications for pharmacists' duty of care. Pharmacists are licensed health professionals whose decisions influence therapy selection, formulary access, patient counseling, and population-level resource allocation. When those decisions rely on numerical claims that do not meet the axioms of measurement, the resulting harm is not speculative; it is foreseeable.

Duty of care in healthcare requires that decisions be grounded in valid evidence and defensible reasoning. This obligation extends beyond clinical pharmacology to the evaluation of therapeutic value, comparative effectiveness, and patient-reported outcomes. When pharmacists are trained to accept utilities, QALYs, and cost-effectiveness ratios as meaningful quantities without understanding that these constructs lack lawful measurement properties, they are placed in an untenable position. They are asked to endorse, communicate, or act upon claims whose numerical form implies precision, trade-offs, and proportionality that do not exist.

The problem is compounded by the authority pharmacists are perceived to hold. Numbers carry persuasive force. When a pharmacist cites an ICER threshold or a QALY gain, patients, clinicians, and decision-makers reasonably assume that these figures represent measurable attributes of health or benefit. In reality, they represent ordinal preferences embedded in simulation narratives. The pharmacist may not intend to mislead, but the professional effect is the same: false certainty is transmitted as evidence.

This creates a breach in professional accountability. A pharmacist cannot meaningfully challenge a formulary exclusion, a restricted access decision, or a pricing justification if they lack the conceptual tools to question whether the underlying metrics are measures at all. Silence becomes complicity. The duty to advocate for patients is weakened because the epistemic basis for advocacy has been hollowed out.

There are also implications for informed consent and patient communication. If benefit claims are derived from non-measures, then discussions of "value," "benefit," or "cost-effectiveness" rest on unstable ground. Patients are entitled to decisions informed by evidence that is not only peer-reviewed but conceptually coherent. Teaching pharmacists to rely on mathematically incoherent constructs undermines that entitlement.

Finally, the failure to teach representational measurement deprives pharmacists of the ability to learn from outcomes. When claims cannot be falsified, errors cannot be corrected. Practice stagnates. The profession becomes administratively compliant rather than scientifically responsive.

In this light, the measurement failures in pharmacy education represent more than curricular shortcomings. They erode the epistemic foundation of professional responsibility. A duty of care that relies on imaginary quantities is not merely weakened; it is structurally compromised.

WHAT A MEASUREMENT-LITERATE DUTY OF CARE WOULD REQUIRE

A measurement-literate duty of care begins with a non-negotiable premise: no numerical claim may be used in decision-making unless it satisfies the axioms required for measurement. This is not a methodological preference but a logical boundary. Pharmacists, as professionals entrusted with evaluating therapy impact and advising on access and use, cannot discharge their duty of care while relying on quantities that are not quantities at all.

First, measurement literacy requires explicit recognition of scale type. Pharmacists must be trained to distinguish nominal, ordinal, interval, and ratio scales and to understand the arithmetic each allows. Ordinal rankings cannot support addition, multiplication, discounting, or aggregation. Any framework that treats preference scores or utilities as interval or ratio measures must therefore be rejected as evidentially invalid. A duty of care requires refusing to act on arithmetic that violates scale properties, regardless of how widely accepted that arithmetic may be.

Second, measurement-literate practice requires unidimensionality as a precondition for any claim about magnitude or change. Composite health state descriptions cannot be measured because they do not represent a single attribute. Teaching pharmacists to accept multiattribute indices as measures undermines their ability to evaluate therapy impact meaningfully. A professional duty of care demands insistence on single-attribute claims or explicit acknowledgment that no measurement has occurred.

Third, where subjective assessment is unavoidable, measurement literacy requires use of Rasch rules. Summed questionnaire scores are not measures. Without Rasch transformation, patient-reported outcomes cannot lawfully be treated as interval quantities, let alone used in comparative or economic evaluation. A pharmacist who understands measurement must recognize that most PRO-based claims fail at this threshold and must treat them accordingly.

Fourth, a measurement-literate duty of care requires falsifiability. Claims about therapy impact must be testable in real time against observable outcomes. Reference-case simulation models do not meet this requirement because they generate narratives conditioned on assumptions rather than measurable predictions. Acting on non-falsifiable claims violates the obligation to base decisions on evidence that can be wrong.

Finally, measurement literacy imposes a responsibility to challenge institutional practices that demand numerical outputs without measurement justification. Compliance is not neutrality. When pharmacists are trained to accept incoherent metrics as professional currency, the duty of care shifts from patient advocacy to administrative convenience.

In sum, a measurement-literate duty of care requires more than technical competence. It requires epistemic discipline: the willingness to say that certain numbers do not mean what they claim to mean, and therefore cannot ethically guide decisions. Without this discipline, the pharmacist's role is reduced from scientific professional to procedural intermediary; precisely what duty of care is meant to prevent.

THE PROFESSIONAL CONSEQUENCES OF CONTINUING

Continuing to practice and teach health technology assessment and pharmacoeconomics as if the measurement critique does not exist carries consequences that extend well beyond academic embarrassment. It erodes professional credibility, distorts clinical judgment, and places pharmacists in an ethically untenable position. The cost of denial is not abstract; it is borne in decision quality, patient trust, and the integrity of the profession itself.

First, there is a loss of epistemic authority. Professions maintain legitimacy by demonstrating that their judgments rest on coherent standards of evidence. When pharmacists participate in decisions grounded in arithmetic applied to non-measures, utilities treated as quantities, QALYs multiplied and aggregated without scale justification, they are no longer exercising scientific judgment. They are executing institutional routines. Over time, this distinction becomes visible to external observers: clinicians, policymakers, and methodologically literate scientists recognize that the numbers cannot support the claims made for them. Once credibility is lost, it cannot be recovered by better models or more data.

Second, there is systematic distortion of clinical and economic reasoning. When false measures are normalized, meaningful distinctions between therapies are blurred or fabricated. Treatments appear cost-effective or not based on artifacts of modeling rather than observable impact. Pharmacists trained under this regime are deprived of the ability to critically interrogate claims, leaving them unable to distinguish evidentiary signal from numerical noise. This undermines their role as safeguards against poor or harmful decisions.

Third, the profession faces an escalating ethical exposure. A duty of care implies acting on evidence that can be defended as meaningful and testable. Continuing to rely on non-falsifiable constructs while knowing, or having reason to know, that "this is how it is done" weakens as critiques become public, systematic, and documented. Silence becomes complicity.

Fourth, there is professional stagnation. Fields that suppress foundational critique lose the capacity to evolve. By treating measurement as settled when it is not, pharmacy education locks future practitioners into a closed methodological loop. Innovation becomes cosmetic: new instruments, new models, new thresholds—all reproducing the same incoherence. The profession risks training graduates for a framework that will eventually be abandoned, leaving them ill-prepared for a transition that could have been gradual and constructive.

Finally, there is a generational cost. Early-career pharmacists inherit a system that demands Before enrolling in a pharmacoeconomics or HTA course, a student should not only ask what will be taught, but why it is being taught and to what end. The central issue is whether the course is aligned with normal science: the cumulative evolution of objective knowledge through testable claims, falsification, and correction. A course that cannot answer this question is not training professionals; it is transmitting convention.

The first question a student should ask is whether the course is explicitly grounded in the evolution of objective knowledge about therapy impact. Normal science proceeds by formulating claims that can be empirically tested, potentially falsified, and revised in light of evidence. If a course treats evaluation as a matter of producing authoritative numbers rather than testing claims, it has already departed from science. This question forces the instructor to clarify whether therapy impact is understood as something to be discovered through measurement and replication, or assumed through accepted constructs.

The second question follows naturally: what theory of measurement does the course use to determine whether claims about therapy impact are even eligible for testing? In normal science, measurement precedes arithmetic. Without lawful measurement, there can be no falsification, because there is nothing stable to test. A course that does not explicitly recognize this relationship cannot contribute to the accumulation of objective knowledge. It trains students to calculate outcomes, not to evaluate claims.

The third question concerns the admissible forms of measurement used to support this scientific process. A defensible course must recognize that only two forms of quantitative measurement can support evaluable HTA claims: linear ratio measures for manifest attributes such as events, time, or resource use, and Rasch logit ratio measures for latent traits such as symptoms or functional capacity. This distinction matters because it defines what can be tested and replicated. Courses that blur this boundary teach students to treat convenience constructs as evidence, undermining the very possibility of scientific progress.

The fourth question addresses how claims are constructed. Normal science advances through single-attribute, unidimensional claims that can be independently evaluated. Students should ask how the course ensures that every value claim meets this standard before being used in decision making. If claims are composite, multidimensional, or rhetorically framed as “overall benefit,” they cannot be falsified. A course that allows such claims is not advancing knowledge; it is producing narratives.

The fifth question connects science to practice: how does the course translate measurement-valid claims into professional competence within health care systems? Pharmacists increasingly serve as formulary managers or formulary committee members. Their value lies not in reproducing calculations, but in their ability to judge whether claims about therapy impact are credible, evaluable, and replicable. A course aligned with normal science equips students to ask the right questions, to reject inadmissible claims, and to insist on evidence that can withstand scrutiny.

This final question exposes the real stakes. Health systems do not need more analysts who can compute ratios. They need professionals who can distinguish evidence from assertion, measurement from scoring, and science from policy preference. A course that emphasizes the evolution of objective knowledge prepares graduates to play that role. It gives them authority grounded in method rather than in precedent.

Taken together, these questions define a course that treats pharmacoeconomics not as a toolkit, but as a scientific discipline embedded in health system decision making. They also define a clear professional benefit. Graduates trained in this way are not locked into defending fragile constructs. They can contribute meaningfully to formulary deliberations by insisting on claims that can be tested, challenged, and improved over time. That is the hallmark of normal science—and the foundation of a defensible professional career.

QUESTIONS A STUDENT SHOULD ASK

Before enrolling in a pharmacoeconomics or HTA course, a student should not only ask what will be taught, but why it is being taught and to what end. The central issue is whether the course is aligned with normal science: the cumulative evolution of objective knowledge through testable claims, falsification, and correction. A course that cannot answer this question is not training professionals; it is transmitting convention.

The first question a student should ask is whether the course is explicitly grounded in the evolution of objective knowledge about therapy impact. Normal science proceeds by formulating claims that can be empirically tested, potentially falsified, and revised in light of evidence. If a course treats evaluation as a matter of producing authoritative numbers rather than testing claims, it has already departed from science. This question forces the instructor to clarify whether therapy impact is understood as something to be discovered through measurement and replication, or assumed through accepted constructs.

The second question follows naturally: what theory of measurement does the course use to determine whether claims about therapy impact are even eligible for testing? In normal science, measurement precedes arithmetic. Without lawful measurement, there can be no falsification, because there is nothing stable to test. A course that does not explicitly recognize this relationship cannot contribute to the accumulation of objective knowledge. It trains students to calculate outcomes, not to evaluate claims.

The third question concerns the admissible forms of measurement used to support this scientific process. A defensible course must recognize that only two forms of quantitative measurement can support evaluable HTA claims: linear ratio measures for manifest attributes such as events, time, or resource use, and Rasch logit ratio measures for latent traits such as symptoms or functional capacity. This distinction matters because it defines what can be tested and replicated. Courses that blur this boundary teach students to treat convenience constructs as evidence, undermining the very possibility of scientific progress.

The fourth question addresses how claims are constructed. Normal science advances through single-attribute, unidimensional claims that can be independently evaluated. Students should ask how the course ensures that every value claim meets this standard before being used in decision making. If claims are composite, multidimensional, or rhetorically framed as “overall benefit,” they cannot be falsified. A course that allows such claims is not advancing knowledge; it is producing narratives.

The fifth question connects science to practice: how does the course translate measurement-valid claims into professional competence within health care systems? Pharmacists increasingly serve as formulary managers or formulary committee members. Their value lies not in reproducing calculations, but in their ability to judge whether claims about therapy impact are credible, evaluable, and replicable. A course aligned with normal science equips students to ask the right questions, to reject inadmissible claims, and to insist on evidence that can withstand scrutiny.

This final question exposes the real stakes. Health systems do not need more analysts who can compute ratios. They need professionals who can distinguish evidence from assertion, measurement from scoring, and science from policy preference. A course that emphasizes the evolution of objective knowledge prepares graduates to play that role. It gives them authority grounded in method rather than in precedent.

Taken together, these questions define a course that treats pharmacoeconomics not as a toolkit, but as a scientific discipline embedded in health system decision making. They also define a clear professional benefit. Graduates trained in this way are not locked into defending fragile constructs. They can contribute meaningfully to formulary deliberations by insisting on claims that can be tested, challenged, and improved over time. That is the hallmark of normal science—and the foundation of a defensible professional career.

ABANDONING AND REPLACING A PHARMACOECONOMICS COURSE

A typical pharmacoeconomics course in schools and colleges of pharmacy follows a now-familiar arc. It opens with the problem of scarcity and the need for efficiency in health care decision making. Students are introduced to costs, outcomes, and the promise of economic evaluation as a rational basis for formulary and access decisions. Numbers appear immediately, long before any discussion of what it means to measure an attribute. Measurement is assumed, not examined.

The course then moves rapidly into cost-effectiveness and cost-utility analysis. Incremental cost-effectiveness ratios are introduced as the central evaluative device. Utilities are presented as numerical representations of health-related quality of life, and QALYs are framed as the natural outcome metric for comparing therapies across diseases. Preference elicitation techniques, time trade-off, standard gamble, visual analogue scales, are taught as alternative ways of "measuring" utility, without any examination of whether these techniques produce measures with the properties required for arithmetic. Ordinal preferences are treated as if they were interval or ratio quantities by convention.

As the course progresses, students learn to manipulate these quantities. They calculate ICERs, interpret cost-effectiveness planes, and apply willingness-to-pay thresholds. Budget impact models and reference-case simulations are introduced as practical tools for decision support. Sensitivity analysis is taught as a way to test robustness, reinforcing the belief that models generate evidence rather than conditional projections. At no point is there a requirement to establish unidimensionality, invariance, or true-zero properties. Arithmetic competence is rewarded; measurement literacy is absent.

Patient-reported outcomes are incorporated as supportive evidence. Students learn to sum Likert scores, calculate mean changes, and apply minimal clinically important differences. These techniques are presented as standard practice. The possibility that ordinal scores cannot support addition, averaging, or multiplication is never raised. Rasch measurement and representational measurement theory are not mentioned. By the end of the course, students are fluent in producing numbers and confident in defending them, even though none of the underlying quantities has been shown to be a measure.

Abandoning this course does not require abandoning economic reasoning. It requires reordering it. A measurement-sound pharmacoeconomics course would begin not with ratios, but with the conditions under which ratios are possible. The opening module would focus explicitly on representational measurement theory: scale types, unidimensionality, invariance, and the requirement that measurement precede arithmetic. Students would learn that not all numbers are measures and that arithmetic operations are licensed only by specific scale properties.

From there, the course would distinguish two admissible classes of quantitative claims. Manifest outcomes—such as events avoided, time to event, hospital days, and resource use—would be treated as single-attribute quantities measured on linear ratio scales. Students would learn how to construct protocol-driven claims around these outcomes, how to test them empirically, and how to replicate them in real-world settings. These claims are limited in scope but scientifically defensible.

Latent traits would be treated differently. If subjective outcomes such as symptom burden or functional status are to be quantified, students would be introduced to Rasch measurement as the only framework capable of producing invariant interval or ratio scales from ordinal responses. They would learn why summated scores are not measures and why many commonly used instruments fail measurement tests. Importantly, they would also learn when not to quantify and how to treat descriptive information as descriptive rather than evidentiary.

Economic reasoning would then be reintroduced on a sound footing. Costs would be linked to single-attribute outcomes rather than composite indices. Models would be taught as exploratory tools, not evidence generators. Students would learn how to separate empirical claims from normative judgments and how to defend that separation professionally.

Such a course would be simpler, not more complex. It would abandon the illusion that everything can be collapsed into a single metric and instead train students to make fewer, clearer, and testable claims. Most importantly, it would align pharmacy education with scientific measurement and protect graduates from having to defend arithmetic that cannot be defended mathematically.

PEAK COUNCIL MEASUREMENT FAILURE

The failure to recognize the critical importance of the axioms of fundamental measurement does not reside primarily with individual pharmacy colleges and schools, nor with their application of HTA as defined within a conventional pharmacoeconomics curriculum. Rather, the source of this failure lies with the two peak educational bodies: the American Association of Colleges of Pharmacy (founded 1900) and the American Foundation for Pharmaceutical Education (founded 1945). These organizations shape the intellectual and educational environment within which pharmacy education operates, influencing curricular norms, research expectations, and the training of successive cohorts of students and researchers. Their role is not peripheral; it is foundational.

The knowledge bases associated with both organizations have been interrogated here and shown to support a consistent pattern of measurement inversion rather than adherence to the axioms required for valid measurement. This is not an isolated or recent development. Over decades, a framework has been institutionalized in which numerical constructs such as utilities, QALYs, and composite endpoints are treated as if they possess the properties necessary for arithmetic operations, despite the absence of demonstrated measurement validity. The consequence is that arithmetic is routinely applied to constructs that do not meet the requirements of unidimensionality, dimensional homogeneity, or scale admissibility.

This sustained omission has profound implications. It suggests that the failure is not due to lack of awareness at the level of individual institutions, but is embedded within the structures that define and transmit knowledge across the discipline. Generations of pharmacy graduates and researchers have been trained within a system that prioritizes methodological convention over measurement validity. Until these peak bodies explicitly recognize and incorporate the axioms of representational measurement into their frameworks, the cycle will continue, and HTA-related claims will remain numerically sophisticated but scientifically unsound.

Waiting for peak bodies to lead this transition is unlikely to produce meaningful change. The historical record suggests that when a framework becomes embedded within institutional guidance, curricula, and professional expectations, it acquires a degree of inertia that resists internal correction. If the adoption of the reference case and its associated constructs has persisted for decades despite the availability of measurement theory, there is little reason to expect that the same bodies will now initiate a fundamental reversal. The implication is clear: the initiative must be taken up by individual colleges and schools, where curricula are designed, teaching priorities are set, and the next generation of practitioners and researchers is formed.

The transition itself is not complex. It does not require the construction of new theoretical systems or the introduction of elaborate methodologies. The essential task is recognition. For manifest attributes, this means restricting claims to linear ratio measures with clearly defined units and true zeros. For latent attributes, it requires the use of Rasch-based models to generate invariant measures expressed on a logit scale. Once these axioms are acknowledged as non-negotiable, the implications follow directly. The current analytical structure, built around the reference case, utilities, QALYs, and composite endpoints, cannot be sustained. It must be set aside, not modified.

What replaces it is a simpler and more disciplined framework. Claims are defined in terms of single attributes, measured according to their scale properties, and evaluated through protocols that allow empirical testing and replication. This is not a retreat from analytical sophistication, but a return to the conditions required for science. The barrier is not technical difficulty, but the willingness to abandon a familiar structure that has never met the standards it claims to uphold.

3. THE TRANSITION TO MEASUREMENT IN HEALTH TECHNOLOGY ASSESSMENT

THE IMPERATIVE OF CHANGE

This analysis has not been undertaken to criticize decisions made by health system, nor to assign responsibility for the analytical frameworks currently used in formulary review. The evidence shows something more fundamental: organizations have been operating within a system that does not permit meaningful evaluation of therapy impact, even when decisions are made carefully, transparently, and in good faith.

The present HTA framework forces health systems to rely on numerical outputs that appear rigorous but cannot be empirically assessed (Table 1). Reference-case models, cost-per-QALY ratios, and composite value claims are presented as decision-support tools, yet they do not satisfy the conditions required for measurement. As a result, committees are asked to deliberate over results that cannot be validated, reproduced, or falsified. This places decision makers in an untenable position: required to choose among therapies without a stable evidentiary foundation.

This is not a failure of expertise, diligence, or clinical judgment. It is a structural failure. The prevailing HTA architecture requires arithmetic before measurement, rather than measurement before arithmetic. Health systems inherit this structure rather than design it. Manufacturers respond to it. Consultants reproduce it. Journals reinforce it. Universities promote it. Over time it has come to appear normal, even inevitable.

Yet the analysis presented in Table 1 demonstrates that this HTA framework cannot support credible falsifiable claims. Where the dependent variable is not a measure, no amount of modeling sophistication can compensate. Uncertainty analysis cannot rescue non-measurement. Transparency cannot repair category error. Consensus cannot convert assumption into evidence.

The consequence is that formulary decisions are based on numerical storytelling rather than testable claims. This undermines confidence, constrains learning, and exposes health systems to growing scrutiny from clinicians, patients, and regulators who expect evidence to mean something more than structured speculation.

The imperative of change therefore does not arise from theory alone. It arises from governance responsibility. A health system cannot sustain long-term stewardship of care if it lacks the ability to distinguish between claims that can be evaluated and claims that cannot. Without that distinction, there is no pathway to improvement; only endless repetition for years to come.

This transition is not about rejecting evidence. It is about restoring evidence to its proper meaning. It requires moving away from composite, model-driven imaginary constructs toward claims that are measurable, unidimensional, and capable of empirical assessment over time. The remainder of this section sets out how that transition can occur in a practical, defensible, and staged manner.

MEANINGFUL THERAPY IMPACT CLAIMS

At the center of the current problem is not data availability, modeling skill, or analytic effort. It is the nature of the claims being advanced. Contemporary HTA has evolved toward increasingly complex frameworks that attempt to compress multiple attributes, clinical effects, patient experience, time, and preferences into single composite outputs. These constructs are then treated as if they were measures. They are not (Table 1).

The complexity of the reference-case framework obscures a simpler truth: meaningful evaluation requires meaningful claims. A claim must state clearly what attribute is being affected, in whom, over what period, and how that attribute is measured. When these conditions are met, evaluation becomes possible. When they are not complexity substitutes for clarity. The current framework is not merely incorrect; it is needlessly elaborate. Reference-case modeling requires dozens of inputs, assumptions, and transformations, yet produces outputs that cannot be empirically verified. Each additional layer of complexity increases opacity while decreasing accountability. Committees are left comparing models rather than assessing outcomes.

In contrast, therapy impact can be expressed through two, and only two, types of legitimate claims. First are claims based on manifest attributes: observable events, durations, or resource units. These include hospitalizations avoided, time to event, days in remission, or resource use. When properly defined and unidimensional, these attributes can be measured on linear ratio scales and evaluated directly.

Second are claims based on latent attributes: symptoms, functioning, need fulfillment, or patient experience. These cannot be observed directly and therefore cannot be scored or summed meaningfully. They require formal measurement through Rasch models to produce invariant logit ratio scales. These two forms of claims are sufficient. They are also far more transparent. Each can be supported by a protocol. Each can be revisited. Each can be reproduced. Most importantly, each can fail. But they cannot be combined. This is the critical distinction. A meaningful claim is one that can be wrong.

Composite constructs such as QALYs do not fail in this sense. They persist regardless of outcome because they are insulated by assumptions. They are recalculated, not refuted. That is why they cannot support learning. The evolution of objective knowledge regarding therapy impact in disease areas is an entirely foreign concept. By re-centering formulary review on single-attribute, measurable claims, health systems regain control of evaluation. Decisions become grounded in observable change rather than modeled narratives. Evidence becomes something that accumulates, rather than something that is re-generated anew for every submission.

THE PATH TO MEANINGFUL MEASUREMENT

Transitioning to meaningful measurement does not require abandoning current processes overnight. It requires reordering them. The essential change is not procedural but conceptual: measurement must become the gatekeeper for arithmetic, not its byproduct.

The first step is formal recognition that not all numerical outputs constitute evidence. Health systems must explicitly distinguish between descriptive analyses and evaluable claims. Numbers that do not meet measurement requirements may inform discussion but cannot anchor decisions.

The second step is restructuring submissions around explicit claims rather than models. Each submission should identify a limited number of therapy impact claims, each defined by attribute, population, timeframe, and comparator. Claims must be unidimensional by design.

Third, each claim must be classified as manifest or latent. This classification determines the admissible measurement standard and prevents inappropriate mixing of scale types.

Fourth, measurement validity must be assessed before any arithmetic is permitted. For manifest claims, this requires confirmation of ratio properties. For latent claims, this requires Rasch-based measurement with demonstrated invariance.

Fifth, claims must be supported by prospective or reproducible protocols. Evidence must be capable of reassessment, not locked within long-horizon simulations designed to frustrate falsification.

Sixth, committees must be supported through targeted training in representational measurement principles, including Rasch fundamentals. Without this capacity, enforcement cannot occur consistently.

Finally, evaluation must be iterative. Claims are not accepted permanently. They are monitored, reproduced, refined, or rejected as evidence accumulates.

These steps do not reduce analytical rigor. They restore it.

TRANSITION REQUIRES TRAINING

A transition to meaningful measurement cannot be achieved through policy alone. It requires a parallel investment in training, because representational measurement theory is not intuitive and has never been part of standard professional education in health technology assessment, pharmacoeconomics, or formulary decision making. For more than forty years, practitioners have been taught to work within frameworks that assume measurement rather than demonstrate it. Reversing that inheritance requires structured learning, not informal exposure.

At the center of this transition is the need to understand why measurement must precede arithmetic. Representational measurement theory establishes the criteria under which numbers can legitimately represent empirical attributes. These criteria are not optional. They determine whether addition, multiplication, aggregation, and comparison are meaningful or merely symbolic. Without this foundation, committees are left evaluating numerical outputs without any principled way to distinguish evidence from numerical storytelling.

Training must therefore begin with scale types and their permissible operations. Linear ratio measurement applies to manifest attributes that possess a true zero and invariant units, such as time, counts, and resource use. Latent attributes, by contrast, cannot be observed directly and cannot be measured through summation or weighting. They require formal construction through a measurement model capable of producing invariant units. This distinction is the conceptual fulcrum of reform, because it determines which claims are admissible and which are not.

For latent trait claims, Rasch measurement provides the only established framework capable of meeting these requirements. Developed in the mid–twentieth century alongside the foundations of modern measurement theory, the Rasch model was explicitly designed to convert subjective observations into linear logit ratio measures. It enforces unidimensionality, tests item invariance, and produces measures that support meaningful comparison across persons, instruments, and time. These properties are not approximations; they are defining conditions of measurement.

Importantly, Rasch assessment is no longer technically burdensome. Dedicated software platforms developed and refined over more than four decades make Rasch analysis accessible, transparent, and auditable. These programs do not merely generate statistics; they explain why items function or fail, how scales behave, and whether a latent attribute has been successfully measured. Measurement becomes demonstrable rather than assumed.

DESIGNED FOR CLOSURE

For those who remain unconvinced that there is any need to abandon a long-standing and widely accepted HTA framework, it is necessary to confront a more fundamental question: why was this system developed and promoted globally in the first place?

The most plausible explanation is administrative rather than scientific. Policy makers were searching for an assessment framework that could be applied under conditions of limited empirical data while still producing a determinate conclusion. Reference-case modeling offered precisely this convenience. By constructing a simulation populated with assumptions, surrogate endpoints, preference weights, and extrapolated time horizons, it became possible to generate a numerical result that could be interpreted as decisive. Once an acceptable cost-effectiveness ratio emerged, the assessment could be declared complete and the pricing decision closed. This structure solved a political and administrative problem. It allowed authorities to claim that decisions were evidence-based without requiring the sustained empirical burden demanded by normal science. There was no requirement to formulate provisional claims and subject them to ongoing falsification. There was no obligation to revisit conclusions as new data emerged. Closure could be achieved at launch, rather than knowledge evolving over the product life cycle.

By contrast, a framework grounded in representational measurement would have imposed a very different obligation. Claims would necessarily be provisional. Measurement would precede arithmetic. Each therapy impact claim would require a defined attribute, a valid scale, a protocol, and the possibility of replication or refutation. Evidence would accumulate rather than conclude. Decisions would remain open to challenge as real-world data emerged. From an administrative standpoint, this was an unreasonable burden. It offered no finality.

The reference-case model avoided this problem entirely. By shifting attention away from whether quantities were measurable and toward whether assumptions were plausible, the framework replaced falsification with acceptability. Debate became internal to the model rather than external to reality. Sensitivity analysis substituted for empirical risk. Arithmetic proceeded without prior demonstration that the objects being manipulated possessed the properties required for arithmetic to be meaningful.

Crucially, this system required no understanding of representational measurement theory. Committees did not need to ask whether utilities were interval or ratio measures, whether latent traits had been measured or merely scored, or whether composite constructs could legitimately be multiplied or aggregated. These questions were never posed because the framework did not require them to be posed. The absence of measurement standards was not an oversight; it was functionally essential.

Once institutionalized, the framework became self-reinforcing. Training programs taught modeling rather than measurement. Guidelines codified practice rather than axioms. Journals reviewed technique rather than admissibility. Over time, arithmetic without measurement became normalized as “good practice,” while challenges grounded in measurement theory were dismissed as theoretical distractions. The result was a global HTA architecture capable of producing numbers, but incapable of producing falsifiable knowledge. Claims could be compared, ranked, and monetized, but not tested in the scientific sense. What evolved was not objective knowledge, but institutional consensus.

This history matters because it explains why the present transition is resisted. Moving to a real measurement framework with single, unidimensional claims does not merely refine existing methods; it dismantles the very mechanism by which closure has been achieved for forty years. It replaces decisiveness with accountability, finality with learning, and numerical plausibility with empirical discipline. Yet that is precisely the transition now required. A system that avoids measurement in order to secure closure cannot support scientific evaluation, cumulative knowledge, or long-term stewardship of healthcare resources. The choice is therefore unavoidable: continue with a framework designed to end debate, or adopt one designed to discover the truth.

Anything else is not assessment at all, but the ritualized manipulation of numbers detached from measurement, falsification, and scientific accountability.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

REFERENCES

---

[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

[2] Krantz D, Luce R, Suppes P, Tversky A. Foundations pf Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

[3] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[4] Wright B. Solving measurement problems with the Rasch Model. *J Educational Measurement.* 1977;14(2):97-116